Statement of Research Intent
Robert Brown

I entered UCLA in 2011 as a PhD student in the bioinformatics department and over the past two and a half years have focused my efforts on developing new methods for analyzing next generation sequencing (NGS) data by leveraging information contained in rare genomic variations. I have worked on three projects directly relating to this interest that have spanned from variant calling from reads, to local ancestry inference and finally to clinical applications with exome sequencing of individuals with Mendelian diseases.

I am preparing to revisit work from my written qualifying exam that used unmapped reads from sequencing data to search for short indel pairs, a situation that occurs when two short (<30 bp) indels occur within one read length of each other. These are known to exist in some cases and statistically are likely to exist much more often than are discovered. I developed a hybrid mapping-assembly approach that accurately detects these regions with a very low false discovery rate.

Work currently in the second round of reviews focuses on performing local ancestry inference of admixed individuals with full genome sequencing data. The method utilizes rare variants with significant population structure to determine variants that are specific to certain population and continental groups. The method significantly reduces run time and is sample aware. With large sample sizes it outperforms current methods. We also show in this manuscript that population specific variants can be used for ancestry inference in sub-continentally admixed individuals, such as an individual with Italian and English ancestry.

My current work is nearing completion and being prepared for publication. It explores how population structure of rare variants can be leveraged to perform more efficient filtering of variants to determine causal loci in Mendelian diseases. We are collaborating with Hane Lee who performs clinical exome sequencing at UCLA. We have used real data she provided to validate our methods and demonstrate that our enhanced filtering methods can significantly reduce the number of candidate causal variants left for prioritization and follow up analysis post-filtering.

As I finish the Mendelian variant filtering project, I am considering looking deeper into the problem of missing heritability. Researchers are increasing looking at rare variation as well as interaction effects to try to find the missing heritability. In addition to using rare variants and interactions, many statistical methods do not consider how variants interact differently depending on if they are on the same or different copies of a chromosome. My work with Mendelian diseases has deepened my understanding of how variants, and combinations of variants, specifically compound heterozygous arrangements, cause detrimental changes to genes. This combined with my local ancestry work has shown me how hidden Markov models could be used to perform an additional haplotype based probabilistic approach to identifying true causal variants that could be masked due to confounding haplotype structure in control individuals. In my future work I would like to explore this possibility with rare variation and its ability to explain missing heritability.

My work has spanned the NGS realm of research through looking at variant calling methods, to academic questions of admixture and population structure of rare variants and is continuing into personalized medicine when looking for causal variants of Mendelian diseases with applications to clinical sequencing. As the cost of sequencing drops and personalized genomic medicine increases in use, academics and doctors will need to be able to quickly and effectively process and analyze sequencing

data. Rare variation will influence many of the clinically important phenotypes, while in academic circles rare variants are helping researchers better understand population structure, migration movements, mutation rates, evolution and much more. In my work I hope to contribute to our understanding of the important role of rare variation.